Crowd Behavior Detection by Statistical Modeling of Motion Patterns

Saira Saleem Pathan, Ayoub Al-Hamadi, Bernd Michaelis
Institute for Electronics,Signal Processing and Communications (IESK),
Otto-von-Guericke-University Magdeburg, Germany
{*Saira.Pathan@ovgu.de, Ayoub.Al-Hamadi@ovgu.de*}

***Abstract*—The governing behaviors of individuals in crowded places offer unique and difficult challenges, and limit the scope of conventional surveillance systems. In this paper, we investigate the crowd behaviors and localize the anomalies due to individual's abrupt dissipation. The novelty of the proposed approach can be described in three aspects. First, we introduce block-clips by sectioning the video segments into non-overlapping spatio-temporal patches to marginalize the arbitrarily complicated and dense flow field. Second, we treat the flow field in each block-clip as 2d distribution of samples and mixtures of Gaussian is used to parameterize it keeping generality of flow field intact. K-means algorithm is employed to initialize the mixture model and is followed by Expectation Maximization for optimization. These mixtures of Gaussian result in the distinct flow patterns precisely a sequence of dynamic patterns for each block-clip. Third, a bank of Conditional Random Field model is employed one for each block clip and is learned from the sequence of dynamic patterns and classifies each block-clip as normal and abnormal. We conduct experiment on two challenging benchmark crowd datasets PETS 2009 and University of Minnesota and results show that our method achieves higher recognition rates in detecting specific and overall crowd behaviors. In addition, the proposed approach shows dominating performance during the comparative analysis with similar approaches in crowd behavior detection.**

***Keywords*-motion analysis; crowd behavior understanding; conditional random field; applications;**

I. Introduction

An attractive aspect of crowd behavior analysis for computer vision is that it allows inference of the self organizing mechanism of dense crowds and supports surveillance operation when it seems to fail. In recent years, crowd behavior analysis is emerged as an entice research with challenging issues of crowd modeling due to jumble of objects, complex scene obstacles, and self-organizing behavior [1].

Behavior analysis is an active research area for both crowded and non-crowded scenes. One natural view of behavior understanding and anomaly detection is that we attempt to analyze components, which naturally characterize the "normal behavior". In contrast, abnormal situations are the deviations from it. Although, the categorization of anomalous activities is usually dependent on the scenarios of interest and treated as a context-sensitive in literature [2] [3].

The problem of behavior analysis can be broken down along several axes. For example, various approaches [4] [5] have been proposed based on modeling the trajectory by tracking each observed object in the scene to characterize the anomalous situations. These approaches show futile performance given the complexity of the problem. Because the crowded scene may contain hundreds or even thousands of objects and operations of object detection, segmentation, tracking is suffered from the problems due to severe object occlusion, varying proximity of objects and similar appearance. Therefore, crowd behavior analysis requires a new set of sophisticated approaches that are being devised by exploiting crowd-specific sociological studies [6].

Several methods have been reported with alternative solutions to avoid above discussed issues particularly tracking. The commonly used features in these solution are optical flow, gradient, spatio-temporal volume to represent the dynamics of crowd. Overall, a diversified literature is available; however, we limit our review to the literature directly relating to the behavior analysis in crowds. Andrade et al. [7] maintains a generative model (i.e. ergodic HMM) at a sparse level for normal motion patterns. Kratz et al. [8] model the statistics of spatio-temporal gradients in cuboids with coupled HMM for dense crowds. Mehran et al. [9] suggested a social force model with the optical flow based particle advection technique and simulate the normal social forces of particles implicitly to detect the deviations from pre-trained parameters.

With same motivation, Wu et al. [10] captures the dynamics of subjects in high density through the tracked trajectories of the advected particles. Albio et al. [11] maintains the probabilities of optical flow at corner points and constitute histograms to detect the deviations and abnormalities on PETS 2009 dataset. In similar context, Benabbas et al. [12] build online probabilistic models of both the density and orientation of flow patterns to detect the crowd activities. Another work is presented by Chan et al. [13] to holistically model the crowd flow in the scene using dynamic texture model where Support Vector Machines (SVM) is used as classifier along with other classifiers to detect the crowd events.

In this paper, we make various levels of contributions to address the problem of modeling and learning the crowd behaviors. We extract the foreground regions and segment video sequence into block-clips. Introduction of our block-clips allows us to concentrate on the issue of representation, specifically how to prototype the dense flow extended over

978-1-4244-7896-5/10/$26.00 ©2010 IEEE

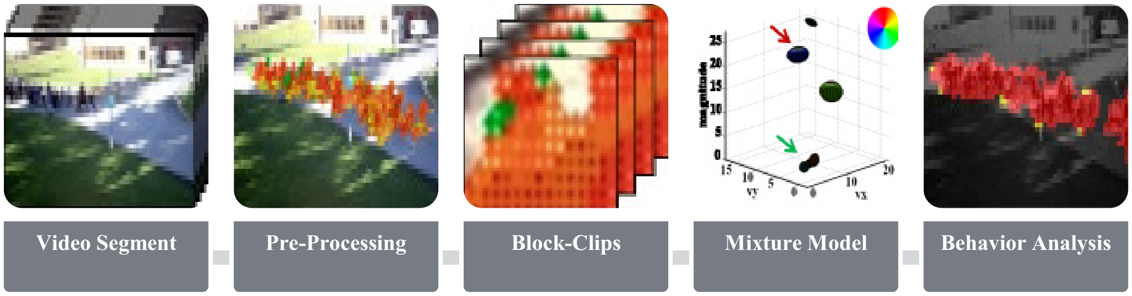

Figure 1. Process flow of the proposed approach

an interval containing significantly correlated and uncorrelated flow field. There on, prior to model crowd dynamics, the prototypes of dense flow representations are parameterized by applying mixtures of Gaussian. K-means clustering is employed to initialize the process and Expectation Maximization (EM) is used to find the maximum likelihood estimators in flow vectors. By doing so, the components with usually low and high variance values especially in flow magnitude can be discarded. In this manner, we are able to handle the optical flow noise in the observed flow field which is not addressed in [7] [9].

The second major contribution is that the dynamics patterns obtained by the mixtures of Gaussian are treated as a sequence of patterns for each block-clip which we termed for simplicity as dynamic patterns or patterns, interchangeably. A bank of Conditional Random Field (CRF) models is constructed, one for each block-clip to model the sequence of dynamic patterns with corresponding label sequence and to characterize the crowd behavior at the specific and global level in an unconstrained environment. In contrast, generative modeling approaches [7] [9](i.e. HMM and LDA) required stringent conditional independence among the observed flow field for more tractable joint distributions. The evaluation is based on the two benchmark datasets by PETS 2009 and UMN whereas the comparative analysis is performed with two related literatures [13] [9] addressing the similar problem.

The organization of this paper is as follows. Section II gives an overview of the proposed approach. In next sections we introduce stages of proposed approach which include pre-processing in section III, mixture model in section IV, behavior detection with CRF in section V. The experimental results are demonstrated in section VI and section VII presents the conclusion of the proposed approach.

II. Overview Of The Proposed Approach

Abrupt activities and complex dynamics collectively define the self-organizing mechanism of the underlying crowded scene. We proposed a top to down approach, which is staged in several phases to model and analyze the characteristics of crowd behaviors as shown in Fig. 1. Our proposed approach begins by extracting the region of interest (ROI) through segmentation. Each video segment is sectioned uniquely into non-overlapping blocks result in spatio-temporal block-clips. The observed samples of flow field inside each block-clip is parameterized by employing the mixtures of Gaussian and constitutes a sequence of dynamic patterns. K-means clustering is used to initialize and to find clusters in this distribution whereas EM is used as an optimization function. The sequence of dynamic patterns for each of our block-clip with corresponding label sequence is used for learning the CRF parameters during training and crowd behaviors are inferred on test samples.

III. Pre-processing

In pre-processing, we build an initial background model generated by using Gaussian Mixture Model (GMM). The foreground is extracted robustly with background subtraction whereas the background model is updated through MDI [14] for each time step (note. currently, we are not handling the problem of shadows).

Each frame is sectioned into N by M blocks of size (i.e. $size = 16$), which is selected after conducting empirical studies over the dataset (i.e. PETS 2009). In parallel, a grid of two by two is placed over the detected ROI which we refer as points of interest (POI) and optical flow is found on these POIs instead of computing on each pixel for example [9]. The flow field at each POI is transformed into a flow vector $f = (v_x, v_y)$. As indicated, v_x and v_y represent the velocities along horizontal and vertical axis of the flow field as shown in Fig. 2(c).

As, it is observed in crowded scenes, the occupancy regions in each frame are equally important and provide distinctive attributes. So, we begin by marginalizing the video sequence into equally sized segments (i.e. video segments) as presented in Fig. 2(a and b). The selection of segment size depends upon the dataset and the frame rate of the video sequence. In our case we kept the size (i.e. $K = 3$) for each video segment s_k. After this, we obtain the block-clips $c_{l,k,t}$ inside each video segments as follows:

$$\mathbf{V} = [s_1, ..., s_K]$$

$$s_k = \left\{c_{(1,1,1)}, ..., c_{(L,K,T)}\right\} \;\; and \quad c_{(l,k,t)} = (f_1, ..., f_P)$$

where $\mathbf{V}$ is the video sequence, s_k is the segmented clip of the sequence containing L block-clips in K-th segment at time t. Each block-clip keeps P cloud of the flow vectors which provides a fundamental information for the analysis of crowd behaviors.

IV. Mixture Model

We define our 2d flow vectors (i.e. $f = (v_x, v_y)$) associated with POI in each block-clip as random samples extended over certain frames (i.e. K) in the corresponding video segment as shown in Fig. 2(c). Since the flow points presented in the block-clip can be significantly different and correlated, therefore it is required to glean the information by applying parametric approximation. Now our objective is to learn and model the component for mixtures of Gaussian for this 2d distribution. Precisely, Gaussian mixture model provides a comprehensive representation of the flow vectors inside each block-clip namely dynamic patterns for block-clips. These dynamic patterns in each block-clip are used to train and test the CRF model for crowd behavior analysis.

Given our 2d distribution of flow vectors in each block-clip, K-means clustering algorithm is employed to initialize the model and to estimate the parameters of these k_c clusters. In practice, each cluster returns a covariance matrix followed by EM, an iterative optimization function for finding the maximum likelihood solutions for our distributions. Particularly, the parameters of k_c clusters in each block-clip is estimated and forms the dynamic patterns as shown in Fig. 2(d). The Gaussian mixture distribution can be written as:

$$p(x) = \sum_{g=1}^{k_c} \pi_g \mathcal{N}(x \,|\mu_g \;, \Sigma_g)$$

where k_c represents the mixtures of Gaussian model, π_g is the weight, μ is the mean and covariance $\sum$ are the parameters of each component of Gaussian model in respective order. The μ_g of the mixtures of Gaussian contains parameters for each dimension of the sample flow vectors (i.e. f). We compute the mean density (d_{μ_g}) for each mixtures of Gaussian thus forming the sequence of dynamic patterns for each block-clip which is to be processed by CRF whereas the length of dynamic patterns sequence (i.e. Seq or $\bar{x}$) is directly proportional to number of clusters k_c. We can write as:

$$\mu_g = (\mu_{g_{vx}}, \mu_{g_{vy}}), \;\; and \;\; d_{\mu_g} = \sqrt{\mu^2_{g_{vx}} + \mu^2_{g_{vy}}}$$

$$Seq = \bar{x} = \left\{d_{\mu_1}, .., d_{\mu_{k_c}}\right\}$$

V. Conditional Random Field and Anomaly Detection

Conditional Random Field is a discriminative modeling technique for labeling the sequential data and a special case of log linear model. CRF provides a probabilistic framework to specify probability of particular label sequence given observation sequence, a very nice description on CRF is presented by Wallach et al. [15]. Particularly, $\bar{x}$ is our input sequence (i.e. $\bar{x} = x_1...x_w$) of w dynamic patterns and $\bar{y}$ is the corresponding label sequence (i.e. $\bar{y} = y_1...y_w$) of respective behaviors. Here, we assume that both sequences $\bar{x}$ and $\bar{y}$ are of same length. As defined by Lafferty et al. [16], the probability of label sequence given observation sequence can be:

$$p(\bar{y}\,|\bar{x}\;;\theta) = \frac{1}{Z(\bar{x},\theta)} \mathtt{exp} \sum_i \theta_i F_i(\bar{x},\bar{y}) \tag{1}$$

The numerator $F_i(\bar{x},\bar{y})$ is the feature function which represents the paired mapping $F_i : X \times Y \longrightarrow \Re$ of the data space X and the label space Y at the different level of granularity. Therefore, feature function F_i can be arbitrarily correlated and defined as follows:

$$F_i(\bar{x},\bar{y}) = \sum_j f_i(y_{j-1}, y_j, \bar{x}, j) \tag{2}$$

where f_i is the low level feature function which is influenced by the subset of the above entities such as, previous label y_{j-1}, current label y_j, observation sequence $\bar{x}$, and current position j.

The denominator in Eq.1 is the partition function commonly termed as normalization factor which ranges over all the label sequence but we assume that the feature-function can depend on at most two labels. So, instead of enumerating all possible $\bar{y}$, this assumption allows us to enumerate the possible $\bar{y}$ efficiently. The formulation of Z is as follows:

$$Z(\bar{x},\theta) = \sum_{\bar{y}} \mathtt{exp} \sum_i \theta_i F_i(\bar{x},\bar{y}) \tag{3}$$

A. Training CRF

We perform training using stochastic gradient methods based on the gradient of conditional likelihood function for nonlinear optimization. The goal of learning task is to compute parameter θ (i.e. weights) values of our model and learns the conditional log-likelihood (CLL) of the training sequences. Therefore, our objective is to maximize the CLL. For this purpose, among many sophisticated techniques, we used stochastic gradient ascent method for training. The formulation is defined in the following:

$$\frac{\partial}{\partial\theta_i \mathtt{log} p(y\,|x\;;\theta)} = F_i(x,y) - \frac{\partial}{\partial\theta_i}\mathtt{log} Z(x,\theta) \tag{4}$$

In the above equation, for each θ_i the partial derivative of CLL is evaluated for a single training sequences (i.e. one wight for each feature-function). Precisely, the partial derivative with respect to θ_i is the i-th value of the feature function for its true label y, minus the averaged feature-function values for all possible labels $\bar{y}$. So, above equation

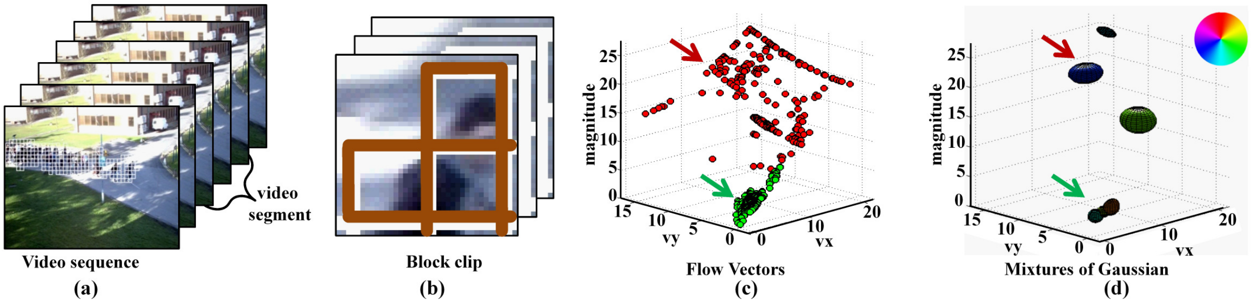

Figure 2. a) presents the sectioning of video sequence into video segments. b) $M \times N$ block-clips are formed in each video-segment. c) demonstrates the observed flow field where green points highlight the normal behavior and red points indicate the abnormal behaviors (Also, arrows indicate the respective crowd behaviors). d) shows the resulting mixtures of Gaussian fitted over the point cloud as shown in (c), whereas colors of the mixtures of Gaussian show the respective orientation of the flow field (please zoom-in for the better visibility).

can be rewritten as:

$$\frac{\partial}{\partial\theta_i}\mathtt{log}p(y\,|x\,;\theta) = F_i(x,y) - E_{\bar{y}\approx p(\bar{y}|x\,;\theta)}\left[F_i(x,\bar{y})\right] \quad (5)$$

In practice, the function $\mathtt{log}(\theta)$ does not maximize in a closed form solution therefore, we invoke BFGS (Broyden Fletcher Goldfarb Shanno) as an optimization routine that estimates the curvature numerically from the first derivative of the CLL and avoid the requirement of exact Hessian inverse computation [17] with stochastic gradient ascent.

B. *Inferencing CRF*

Given the test sequence of dynamic patterns for each block-clip $\bar{x}$ and learned parameter values of θ from the training data, the corresponding label for the sequence is obtained as:

$$\bar{y^*} = \mathtt{argmax}_{\bar{y}} p(\bar{y}\,|\bar{x}\,;\theta) = \mathtt{argmax}_{\bar{y}}\sum_i \theta_i F_i(\bar{x},\bar{y}) \quad (6)$$

Using the definition of feature function in Eq.2, we get:

$$\bar{y^*} = \mathtt{argmax}_{\bar{y}}\sum_i \theta_i \sum_j f_i(y_{j-1},y_j,\bar{x},j) \quad (7)$$

Each label sequence is aggrandize from $< start, end >$ states of labels (i.e. y_0 to y_{n+1}), so for efficient computation an alternative choice is to employ matrices. For this, g_j is a $q \times q$ matrix where q is the cardinality of the set vectors in the label sequence $\bar{y}$ and is defined over each pair of labels y_{j-1} and y_j as follows:

$$g_j(y_{j-1},y_j\,|\bar{x}) = \mathtt{exp}(\sum_i \theta_i f_i(y_{j-1},y_j,\bar{x},j)) \quad (8)$$

So, for each j, we will get different g_j function which depends on weight θ, test observation sequence $\bar{x}$ and the position j. The sequence probability of the label $\bar{y}$ given observation sequence $\bar{x}$ can be rewritten in compact manner in the following:

$$p(\bar{y}\,|\bar{x}\,;\theta) = \frac{1}{Z(\bar{x},\theta)}\prod_j g_j(y_{j-1},y_j\,|\bar{x}) \quad (9)$$

$$Z(\bar{x},\theta) = \prod_j g_j(y_{j-1},y_j) \quad (10)$$

Our main contention in obtaining the local sequence of dynamic patterns in each block-clip is that when global information (i.e. at video segment level) of flow field is used, it is difficult to reveal the required level of detail which can differentiate coherent and incoherent dynamics. Therefore, in the above, we obtained the intrinsic motion flow features in a compact manner that faithfully characterizes the behavior of the crowd dynamics. The GMM is invoked to parameterize dense flow vectors into a sequence of dynamic patterns which are modeled with CRFs to characterize the normal and abnormal behaviors in the crowd physics.

VI. Experiments and Discussion

The proposed approach is tested on two publicly available benchmark datasets from PETS 2009 [18] and UMN [19]. Ideally, normal situation is represented by the usual walk of large number of people whereas the corresponding abnormal situations (i.e. running, panic and dispersion) are observed when individuals or group of individuals deviate from the normal behavior. There is a major distinction between these two datasets, for example, in PETS 2009, the abnormality begins gradually unlike UMN dataset, which makes PETS more challenging due to the transitions from normal to abnormal situations. Table. 1 indicates the scenarios and the datasets used for the training process (Note. UMN dataset is tested without additional training).

Table I
Training Process

training scenario	training set	training frames
S1.L1	13-57	220
S1.L1	13-59	240
S1.L2	14-06	200
S1.L3	14-17	90
S1.L3	14-33	343

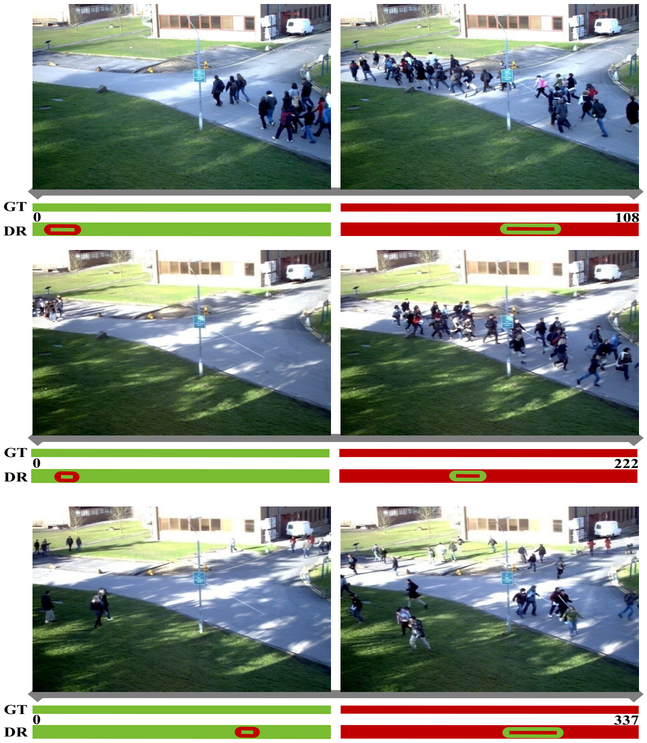

Figure 3. presents the quantitative analysis on PETS 2009. The left frames indicate absolute normal (green) behavior and the right frames depict absolute abnormal (red) behaviors along the time-line (gray) in each row.

A. Qualitative Analysis

A qualitative presentation indicates the ground truth (GT) and the detection result (DR) in each row for normal and abnormal situations in the sequences as shown in Fig. 3. The right and left frame in each sequence depicts the normal and abnormal behavior of the crowd where the thin color bars (i.e. green and red) in each rows show the ground truth and thick bars indicate the detection results. The colors of the bars define the crowd behaviors and timings of the occurrences. The incorrect localization of the crowd behaviors are marked with respective colors of false detections in Fig. 3.

Fig. 4 and Fig. 5 demonstrate the detection results of crowd behaviors on PETS 2009 and UMN datasets. Normal behaviors are marked as green and abnormal behaviors are highlighted with red in frames. The results show that the proposed approach is capable of locating the specific and overall abnormalities in the regions that are occupied by the crowd.

B. Quantitative Analysis

Table. II shows the confusion matrix of the probability of normal and abnormal behavior analysis of crowd dynamics for each class. The diagonal elements in the confusion matrices represent the percentage probability of each class in the group. Misclassification between the classes are shown by the non-diagonal elements which are observed due to prominent motion field at objects legs parts as compared to body and head of the objects.

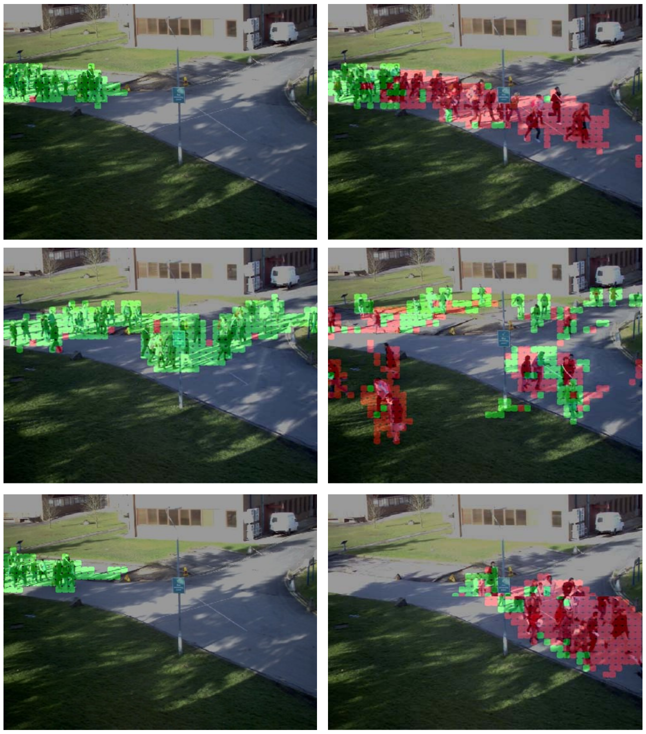

Figure 4. shows the detection results on PETS 2009. Left frames show the normal behavior detection indicated by green patches and right frames depict the abnormal behaviors marked with red patches.

Table II
NORMAL AND ABNORMAL BEHAVIOR DETECTION

Event	Normal	Abnormal
Normal	97.8	2.2
Abnormal	3.1	96.9

To analyze the performance of our proposed approach in detecting the crowd dynamics effectively, we have made comparative analysis from two recent proposed techniques [9] [13]. In the first approach, the computed social forces are model with LDA whereas in the second approach SVM is used to classify the behaviors (we considered the categories of behaviors which define the abnormality). As can be seen in Table. III, the performance of our method is promising and achieving higher detection rate in localizing the crowd behaviors when compared with related approaches.

VII. CONCLUSION

We propose a novel approach for detecting crowd behaviors by modeling computed sequence of dynamic patterns using Conditional Random Field. We define block-clips as non-overlapping spatio-temporal patches and parameterize the flow vectors in each block-clip with mixtures of Gaussian

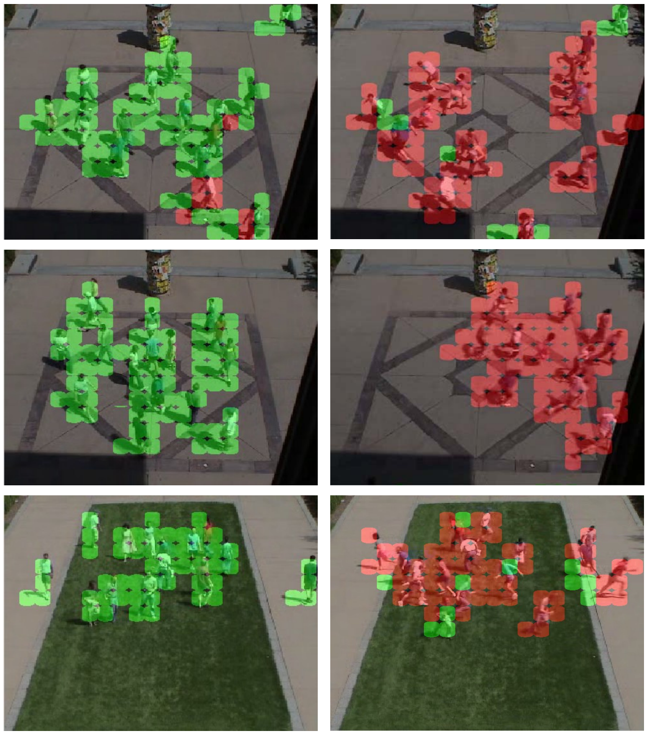

Figure 5. shows the detection results on UMN. Left frames show the normal behavior detection indicated by green patches and right frames depict the abnormal behaviors marked with red patches.

Table III
COMPARATIVE ANALYSIS

Methods	Results(%)
Our Method	97.3
Mehran et al.[9]	96
Chen et al.[13]	81

to obtain a sequence of dynamic patterns. The sequence of dynamic patterns for each block-clip allows an effective representation of features and CRF is learned from these sequences to characterize the behaviors. The results of our method, indicates that the proposed approach is effective in detection and localization of specific and overall behaviors in the crowd. The presented results show promising performance and outperform when compared with the related work.

ACKNOWLEDGMENT

This work is supported by Forschungspraemie (BMBF:FKZ 03FPB00213) and Transregional Collaborative Research Centre SFB/TRR 62 funded by the German Research Foundation (DFG).

REFERENCES

[1] B. Zhan, D. Monekosso, P. Remagnino, S. Velastin, and L. Xu, "Crowd analysis: A survey," *Machine Vision Application*, vol. 19, no. 5-6, pp. 345–357, 2008.

[2] T. Xiang and S. Gong, "Online video behaviour abnormality detection using reliability measure," in *British Machine Vision Conference*, 2005.

[3] H. Dee and D. Hogg, "Detecting inexplicable behaviour," in *In Proceedings of the British Machine Vision Conference, The British Machine Vision Association*, 2004, pp. 477–486.

[4] G. Dalley, X. Wang, and E. Grimson, "Event detection using an attention-based tracker," in *Proceedings of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance in Conjunction with ICCV*, 2007, pp. 71–78.

[5] C. Piciarelli, C. Micheloni, G. L. Foresti, and S. Member, "Trajectory-based anomalous event detection," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, pp. 1544–1554, 2008.

[6] A. Johansson, "Constant-net-time headway as a key mechanism behind pedestrian flow dynamics," *Physical Review E*, vol. 80, no. 2, pp. 026 120–1, Aug 2009.

[7] E. L. Andrade, B. Scott, and R. B. Fisher, "Hidden markov models for optical flow analysis in crowds," in *Proceedings of International Conference on Pattern Recognition*. IEEE Computer Society, 2006, pp. 460–463.

[8] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[9] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 935–942, 2009.

[10] S. Wu, B. E. Moore, and M. Shah, "Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[11] A. Albiol, M. Silla, A. Albiol, and J. Mossi, "Video analysis using corner motion statistics," in *Performance Evaluation of Tracking and Surveillance Workshop at CVPR*, 2009, pp. 31–37.

[12] Y. Benabbas, N. Ihaddadene, and C. Djeraba, "Global analysis of motion vectors for event detection in crowd scenes," in *Performance Evaluation of Tracking and Surveillance workshop at CVPR*, 2009, pp. 109–116.

[13] A. B. Chan, M. Morrow, and N. Vasconcelos, "Analysis of crowded scenes using holistic properties," in *Performance Evaluation of Tracking and Surveillance workshop at CVPR*, 2009, pp. 31–37.

[14] A. Al-Hamadi and B. Michaelis, "An intelligent paradigm for multi-objects tracking in crowded environment," *Journal of Digital Information Management*, vol. 4, pp. 183–190, 2006.

[15] H. M. Wallach, "Conditional random fields: An introduction," University of Pennsylvania, Tech. Rep. MS-CIS-04-21, 2004.

[16] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," *In Proceedings of International Conference on Machine Learning*, pp. 282–289, 2001.

[17] F. Sha and F. Pereira, "Shallow parsing with conditional random fields," 2003.

[18] J. Ferryman and A. Shahrokni, "PETS2009," $www.cvg.rdg.ac.uk/PETS$2009.

[19] UMN, "Detection of unusual crowd activity," $http : //mha.cs.umn.edu$.